

Sistema automatizado para monitorear el uso de equipos de protección personal en la industria de la construcción

Massiris, M.^{a,b,*}, Fernández, J. A.^c, Bajo, J.^a, Delrieux, C.^{a,b}

^a Departamento Ingeniería Eléctrica y de Computadoras, Universidad Nacional del Sur, Av. San Andrés, n°800, 8000, Bahía Blanca, Argentina.

^b Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina (CONICET), Av. La Carrindanga, km. 7, 8000, Bahía Blanca, Argentina.

^c Departamento Ingeniería Eléctrica, Electrónica y Automática, Universidad de Extremadura, Av. Elvas, s/n, 06006, Badajoz, España.

To cite this article: Massiris Fernández, M., Fernández Muñoz, J. A., Bajo, J., Delrieux, C. 2021. An automated system for monitoring the use of personal protective equipment in the construction industry. Revista Iberoamericana de Automática e Informática Industrial 18, 68-74. <https://doi.org/10.4995/riai.2020.13243>

Resumen

Este artículo presenta un novedoso sistema de visión por computador que genera indicadores automatizados de uso adecuado de equipos de protección personal (EPP) de gran importancia en la industria de la construcción, concretamente el uso de casco de seguridad y chaleco de alta visibilidad. El sistema se construye sobre una arquitectura de redes neuronales que trabaja sobre imágenes digitales. Primero se utiliza la red neuronal OpenPose para la detección de puntos antropométricos de los trabajadores visualizados, los cuales se utilizan para segmentar automáticamente regiones de interés (ROI) ubicadas en la cabeza y el tronco. Sobre estas ROI, un clasificador neuronal estima la presencia o ausencia de los dos EPP de interés. Los resultados obtenidos en vídeos tomados en movimiento por drones o *smartphones* muestran que nuestro sistema es plenamente capaz de realizar una evaluación completa de indicadores de utilización de estos dos EPP sin asistencia, con el propósito principal de prevenir incidentes potencialmente peligrosos en el lugar de trabajo.

Palabras clave: Automatización, prevención de riesgos laborales, equipo de protección personal, redes neuronales, visión por computador.

An automated system for monitoring the use of personal protective equipment in the construction industry

Abstract

We present a novel computer vision system which generates automated indicators of proper use of personal protective equipment (PPE) of great importance in the construction industry, specifically the use of safety helmet and high visibility vest. The system is built on a neural network architecture that works on digital images. First, the OpenPose network is used for the detection of anthropometric points of the visualized workers. These points are used next to automatically segment regions of interest (ROI) located about a worker's head and trunk. On these ROIs, a neuronal classifier estimates the presence or absence of each PPE of interest. Obtained results in moving videos from drones or smartphones show that our system is fully capable of carrying out a complete evaluation of usage indicators of these two PPEs without human intervention, with the main purpose of preventing potentially dangerous incidents in the workplace.

Keywords: Automation, occupational risk prevention, personal protective equipment, neural networks, computer vision.

*Autor para correspondencia: manlio.massiris@uns.edu.ar

1. Introducción

La Directiva 89/686/EEC de la Unión Europea (1989) estipula que los equipos de protección personal (EPP) siempre deben ser utilizados cuando los riesgos laborales no puedan evitarse o limitarse suficientemente por medios técnicos de protección colectiva o mediante medidas, métodos o procedimientos de organización del trabajo. Por ello, los EPP son la última barrera entre el trabajador y el riesgo (ILO, 2002).

A pesar de estas regulaciones, la realidad muestra que en muchos sitios de construcción no se utilizan adecuadamente los EPP, tanto por parte de los trabajadores como de los visitantes ocasionales. Los informes públicos de seguridad y salud laboral europeos y norteamericanos continúan indicando que la mayoría de lesiones laborales producidas por impacto en la cabeza se deben a la falta de uso del casco de seguridad (Park et al., 2015). Esto suele asociarse a una combinación de factores, destacando entre ellos la incomodidad de uso, asociada frecuentemente a una obstaculización de la realización normal del trabajo, la baja promoción de los hábitos de uso por parte del contratista y, en menor medida, los costes asociados (ILO, 2002; Arias Gallegos, 2011).

Actualmente, el proceso de monitoreo del uso de EPP se realiza principalmente mediante observación *in situ* por parte de expertos, lo que resulta costoso por la necesidad de ocupar personal especializado para el desarrollo de esta tarea. Por otra parte, este método habitual presenta una notable variabilidad en los resultados obtenidos por diferentes evaluadores, e incluso por un mismo evaluador en diferentes contextos. La variabilidad *inter*-observador se debe principalmente a factores subjetivos como la formación, experiencia o medios de observación, mientras que la variabilidad *intra*-observador se asocia normalmente a la fatiga cognitiva y a cambios en el entorno de observación (Park et al., 2015; Massiris Fernández et al., 2018).

Ante estas dificultades, los recientes avances en tecnologías de Visión por Computador (CV, en inglés) y aprendizaje automático están modificando el estado del arte del monitoreo de utilización de EPP. En la última década, la detección de EPP ha sido utilizada como filtro inicial en procesos de seguimiento visual de trabajadores en sitios de construcción. Este filtro es un clasificador de personas trabajadoras basado en la hipótesis de que todo trabajador objetivo debe utilizar ropa de alta visibilidad (Park and Brilakis, 2016; Seong et al., 2017; Mosberger et al., 2014; Konstantinou et al., 2019). La clasificación de estos EPP se materializa mediante detectores de color, que pueden diseñarse en distintos modelos como RGB, HSV o Lab, para determinar en una imagen la existencia de objetos con colores de seguridad conocidos (Park and Brilakis, 2016; Seong et al., 2017; Konstantinou et al., 2019; Mosberger et al., 2014).

Alternativamente, se han utilizado distintas combinaciones de otros descriptores visuales para construir detectores de EPP, p. ej. descriptores basados en la forma, la presencia de bordes o convexidades (Shrestha et al., 2015; Mneymneh et al., 2017), o histogramas de gradientes orientados (HOG, en inglés) (Memarzadeh et al., 2013; Park and Brilakis, 2012). Estos descriptores pueden agruparse para formar atributos asociados a modelos que pueden procesarse mediante técnicas de aprendizaje automático y clasificación de patrones como Redes Neuronales

Artificiales (ANN, en inglés) y Máquinas de Vectores de Soporte (SVM, en inglés) (Park et al., 2017; Brilakis et al., 2011).

Sin embargo, aunque el reconocimiento de EPP basado en descriptores visuales puede funcionar bien en situaciones controladas, en el marco operativo de la industria de la construcción, el entorno de trabajo habitual genera importantes dificultades (Xie et al., 2018). En imágenes tomadas en exteriores, los descriptores visuales asociados a EPP pueden sufrir alteraciones relevantes debidas a factores no evitables como la postura del trabajador respecto al punto de observación, la iluminación del entorno de trabajo y las oclusiones visuales debidas a otros trabajadores y sus herramientas de trabajo, entre otros (Seong et al., 2017; Mosberger et al., 2014). Esta situación es similar a la que ocurre en muchos otros contextos donde tradicionalmente se ha aplicado la ingeniería de atributos para entrenar sistemas de CV para propósitos diversos. Si se desea un método de clasificación confiable, el enfoque debe ser robusto frente a estas variaciones. Por dicho motivo, los trabajos más recientes sobre detección visual de EPP exploran técnicas de aprendizaje profundo sobre ANN multicapa para lograr un mejor rendimiento en el reconocimiento visual de patrones en entornos complejos y de alta variabilidad (Massiris Fernández et al., 2018; Fang et al., 2018; Xie et al., 2018; Son et al., 2019), siendo esta metodología la que se adopta en este trabajo.

2. Materiales y métodos

2.1. Metodología

Nuestro objetivo es diseñar y probar una metodología para la detección de la correcta utilización de EPP en videos sin cámara fija, a través de técnicas de CV y aprendizaje automático, para prevenir accidentes laborales. El Algoritmo 1 y la Figura 1 presentan una descripción general de la metodología adoptada. Cada cuadro del video de entrada *Vid* se analiza inicialmente con una red OpenPose (OP). OP destaca respecto a otras ANN del estado del arte por una elevada tasa de detección visual de personas mediante la identificación de su esqueleto (Cao et al., 2017). Se ha seleccionado para este trabajo la versión más pequeña, OP18, ya que sus esqueletos contienen 18 puntos antropométricos, suficientes para nuestro propósito.

OP18 toma una imagen digital RGB como entrada a una ANN multicapa de dos etapas. En la primera etapa se predicen los mapas de confianza de la detección de puntos antropométricos del cuerpo, mientras que en la segunda se asocian dichos puntos con todos los posibles esqueletos. OP18 obtiene a su salida un conjunto variable de tuplas ($E.x, E.y, E.c$), asociadas a la posición estimada 2D (x, y) y la confiabilidad de la estimación, c , de 18 puntos antropométricos agrupados en esqueletos E según la Figura 2.

La confiabilidad de cada punto OP es una cantidad numérica en el rango $[0, 1]$ que depende de aspectos tales como resolución, distorsión proyectiva y oclusiones, debidas estas últimas tanto a la existencia de elementos opacos en la escena, como a la postura relativa del trabajador respecto a la cámara (Xie et al., 2018). Para descartar detecciones de baja certeza, se definen umbrales de tamaño y confiabilidad mínimos para esqueletos, u_{Tam}^E y u_{Conf}^E , asociados respectivamente al tamaño $E_{Tam} = [\max(E.x) - \min(E.x), \max(E.y) - \min(E.y)]$ y confiabilidad $E_{Conf} = \frac{1}{18} \sum_{j=0}^{17} E.c[j]$ del esqueleto E .

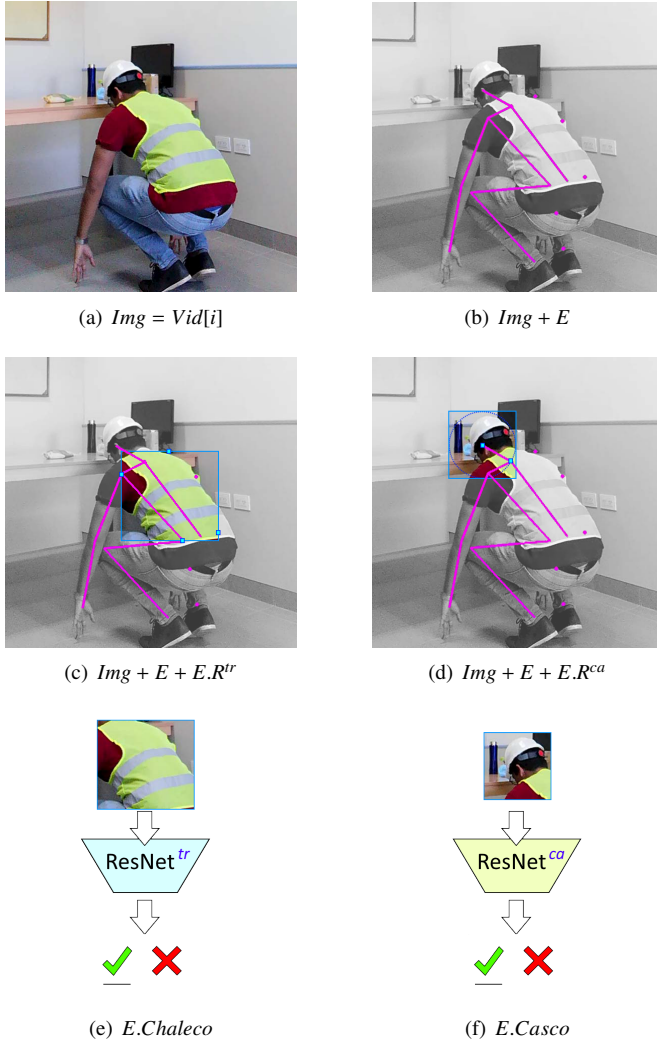


Figura 1: Resumen del método desarrollado.

Algoritmo 1: Detección antropométrica de EPP

Input: Vid , u_{Conf}^E , u_{Tam}^E , u_{Conf}^{ca} , u_{Tam}^{ca} , u_{Conf}^{tr} , u_{Tam}^{tr} ;

forall i **in** Vid **do**

$img \leftarrow Vid[i]$;

$esqueletos \leftarrow OP(img)$;

forall k **in** $esqueletos$ **do**

$E \leftarrow esqueletos[k]$;

if $E_{Conf} > u_{Conf}^E$ **and** $E_{Tam} > u_{Tam}^E$ **then**

$E.R^{ca} \leftarrow segmentarCabeza(E)$;

if $E.R^{ca}_{Conf} > u_{Conf}^{ca}$ **and** $E.R^{ca}_{Tam} > u_{Tam}^{ca}$ **then**

$E.Casco \leftarrow RN^{ca}(img[E.R^{ca}])$

end

$E.R^{tr} \leftarrow segmentarTronco(E)$;

if $E.R^{tr}_{Conf} > u_{Conf}^{tr}$ **and** $E.R^{tr}_{Tam} > u_{Tam}^{tr}$ **then**

$E.Chaleco \leftarrow RN^{tr}(img[E.R^{tr}])$

end

end

end

Result: $E[i][k].Casco$, $E[i][k].Chaleco$

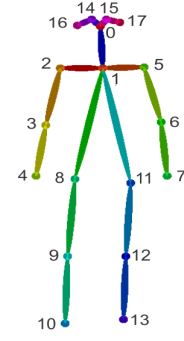


Figura 2: Puntos antropométricos en OpenPose (OP) (Cao et al., 2017).

Para cada E obtenido con tamaño E_{Tam} y confiabilidad E_{Conf} suficientes, se determina una región de interés (ROI) asociada a la cabeza, $E.R^{ca}$, y otra asociada al tronco, $E.R^{tr}$, donde se espera encontrar los EPP de interés. Como se muestra en la Figura 1, estas ROI se definen a partir de puntos OP relacionados con dichas partes del cuerpo, con índices $j_{ca} = [1, 16, 17]$ para $E.R^{ca}$ y $j_{tr} = [1, 2, 5, 8, 11]$ para $E.R^{tr}$. Esto permite definir, al igual que para E , valores de tamaño y confiabilidad mínimos para cada ROI, denominados u_{Conf}^{ca} , u_{Tam}^{ca} , u_{Conf}^{tr} y u_{Tam}^{tr} en el Algoritmo 1.

Finalmente, las subimágenes obtenidas a partir de $E.R^{ca}$ y $E.R^{tr}$ con tamaño y confiabilidad suficientes, son procesadas por redes ResNet (RN) entrenadas para detectar la presencia de un EPP específico. ResNet (*Residual Network*) es una ANN de tipo convolucional que se caracteriza por la inclusión de conexiones residuales, que permiten extender la conectividad directa de una capa a cualquier capa posterior, lo que facilita aumentar el número de capas sin perder capacidad de entrenamiento (He et al., 2016). Esto las convierte en una de las mejores redes de clasificación en términos de precisión frente a coste computacional. En este trabajo se ha elegido el modelo ResNet de 18 capas, RN18. Para adaptar esta red a nuestra problemática, se re-entrenaron las tres últimas capas de cada red RN18 para identificar cada EPP objetivo.

Por tanto, las redes clasificadoras RN18^{ca} y RN18^{tr} asignan respectivamente a cada E dos etiquetas, $E.Casco$ y $E.Chaleco$, ambas con valores (Sí, No). Por último, cada E procesado se almacena en una base de datos para su posterior tratamiento estadístico.

2.2. Umbrales y ROIs

Para el propósito de esta investigación, la función *segmentarTronco()* del Algoritmo 1 acota $E.R^{tr}$ al rectángulo de área mínima, con lados paralelos a los bordes de la imagen, que contiene todos los puntos antropométricos de E asociados a la cintura y los hombros (véase Figura 1(c)). Por consiguiente, su tamaño y confiabilidad se definen como $E.R^{tr}_{Tam} = [\max(E.x[j_{tr}]) - \min(E.x[j_{tr}]), \max(E.y[j_{tr}]) - \min(E.y[j_{tr}])]$ y $E_{Conf} = \frac{1}{5} \sum E.c[j_{tr}]$, con $j_{tr} = [1, 2, 5, 8, 11]$.

Para definir la región de la cabeza de un esqueleto, los puntos antropométricos de referencia más comunes son la base del cuello y el punto medio entre los oídos (Ankrum and Nemeth, 2000). Sin embargo, ambos oídos solo son detectables en los planos frontal y posterior, mientras que en los planos laterales solo hay un oído visible. Por esta razón, la función

segmentarCabeza() del Algoritmo 1 delimita $E.R^{ca}$ como un cuadrado con lados paralelos a los bordes de la imagen, cuyo centroide ($E.R_{Cent}^{ca}$) se calcula como sigue. Si en un esqueleto se detectan ambos oídos, se toma el centroide como el punto medio entre ellos: $E.R_{Cent}^{ca} = \frac{1}{2}(E[16] + E[17])$. Por el contrario, si un único oído es detectado, el centroide es asignado a este: $E.R_{Cent}^{ca} = E[16]$ ó $E[17]$. Por último, el tamaño de la ROI de la cabeza se establece como el doble de la distancia existente entre la base del cuello y el mencionado centroide $E.R_{Tam}^{ca} = [2r, 2r]$, con $r = \|E.R_{Cent}^{ca} - E[1]\|_2$ (véase Figura 1(d)).

Según lo establecido en el apartado anterior, los umbrales de tamaño y confiabilidad para el esqueleto (u_{Tam}^E y u_{Conf}^E), la cabeza (u_{Tam}^{ca} y u_{Conf}^{ca}) y el tronco (u_{Tam}^{tr} y u_{Conf}^{tr}) se utilizan para descartar esqueletos y ROIs objetivo que puedan presentar problemas en la etapa de clasificación. En particular, los umbrales de confiabilidad evitan considerar resultados poco precisos, mientras que los umbrales de tamaño filtran pequeñas ROIs que pueden generar clasificaciones falibles. En los experimentos del presente trabajo se utilizan los siguientes umbrales de confianza: $u_{Conf}^E = 0,35$, $u_{Conf}^{ca} = 0,3$ y $u_{Conf}^{tr} = 0,3$. Para los esqueletos se define $u_{Tam}^E = 0,2$ como valor mínimo de la relación entre la altura del esqueleto E_{Tam} y la altura del fotograma. Finalmente, dado que los clasificadores RN18 están diseñados para recibir a su entrada una imagen RGB de 224×224 px, se utiliza un umbral mínimo sobre la longitud de la diagonal de cada ROI de interés $u_{Tam}^{ca} = 112$ px y $u_{Tam}^{tr} = 112$ px.

2.3. Dataset y entrenamiento

El dataset construido para este trabajo está formado por ROIs extraídas como se indica en la Sec. 2.2 de más de 30 vídeos disponibles en Internet, que recogen diversas actividades reales de trabajadores de la industria de la construcción, con distintas condiciones de iluminación y perspectiva de escena. Para acelerar el proceso de etiquetado, se seleccionaron vídeos de secuencia continua donde todos los trabajadores visualizados pueden clasificarse directamente con las mismas etiquetas de uso de los EPP objetivo en cualquier instante de tiempo.

Posteriormente, en la selección de fotogramas se ha maximizado la diversidad en las muestras. Así, para las clases de clasificación positiva de *Casco* y *Chaleco*, se incluyeron varios colores y modelos. En la clase negativa de *Casco* se incluyeron trabajadores sin cabello, con distintos cortes y colores de cabello, con y sin gorras o sombreros distintos de un casco de seguridad homologado (véase Figura 3). Finalmente, en la clase negativa de *Chaleco* se incluyeron torsos desnudos o con camisetas y camisas de distintas tramas y colores, así como variabilidad de géneros (véase Figura 4). En definitiva, el dataset consta de un total de 23.303 subimágenes de cabezas y 21.378 subimágenes de torsos, obtenidas de esqueletos asociados a trabajadores de la industria de la construcción.

Para el entrenamiento del clasificador RN18^{ca}, se realizó una partición aleatoria con 15.367 subimágenes $E.R^{ca}$ para re-entrenamiento y 7.666 para validación, mientras que para el clasificador RN18^{tr}, se eligieron de igual forma 14.264 subimágenes $E.R^{tr}$ para re-entrenamiento y 7.114 para validación. Finalmente, ambos conjuntos de re-entrenamiento se dividieron en un 80%-20% para entrenamiento y testeo, respectivamente, proporcionando los resultados recogidos en las matrices de confusión de las Tablas 1 y 2, que

contienen las definiciones usuales de parámetros de rendimiento de clasificación: sensibilidad = $vp/(vp + fn)$, precisión = $vp/(vp + fp)$, especificidad = $vn/(vn + fp)$, valor predictivo negativo (NPV) = $vn/(vn + fn)$ y exactitud = $(vp + vn)/(vp + vn + fp + fn)$.

Tabla 1: Matriz de confusión para la validación de RN18^{ca} (*Casco de seguridad*).

| | | Real | | Total |
|----------|---|----------------------|-----------------------|----------------------------|
| | | v | f | |
| Predicho | p | $vp = 4.546$ | $fp = 12$ | 4.558 Precisión = 0,997 |
| | n | $fn = 13$ | $vn = 3.095$ | 3.108 NPV = 0,995 |
| Total | | 4.559 | 3.107 | 7.666 |
| | | Sensibilidad = 0,997 | Especificidad = 0,996 | Exactitud = 0,996 |

Tabla 2: Matriz de confusión para la validación de RN18^{tr} (*Chaleco de alta visibilidad*).

| | | Real | | Total |
|----------|---|----------------------|-----------------------|----------------------------|
| | | v | f | |
| Predicho | p | $vp = 2.754$ | $fp = 22$ | 2.776 Precisión = 0,992 |
| | n | $fn = 18$ | $vn = 4.320$ | 4.338 NPV = 0,996 |
| Total | | 2.772 | 4.342 | 7.114 |
| | | Sensibilidad = 0,994 | Especificidad = 0,995 | Exactitud = 0,994 |

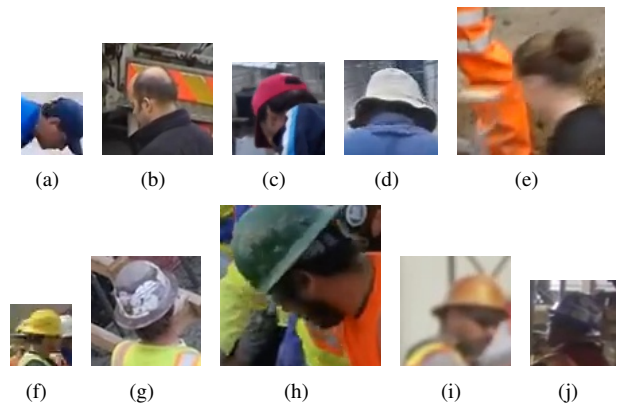


Figura 3: Dataset $E.R^{ca}$ con etiquetas de “Casco” (a)-(e) No y (f)-(j) Sí.



Figura 4: Dataset $E.R^{tr}$ con etiquetas de “Chaleco” (a)-(f) No y (g)-(l) Sí.

El entrenamiento de las redes RN18 se realizó en la plataforma de computación en la nube Google Colab utilizando el paquete Fast.ai, mientras que OP18 se instaló de forma local con Pytorch en un PC con procesador Intel Core i7 con 16 GB de RAM y tarjeta gráfica NVIDIA GeForce GTX 1080 TI. El software de evaluación se diseñó para ejecución local en GPU en lenguaje Python sobre Ubuntu de las tres ANN, alcanzando una velocidad media de procesamiento de 3, 57 fps.

3. Resultados experimentales

Para la validación de la metodología propuesta se seleccionaron tres vídeos de actividades laborales típicas del sector, que presentan diferentes condicionantes técnicos: (1) cámara estabilizada en dron volando a baja altura a mediodía (Figura 5), (2) cámara deportiva fijada a la cabeza de un trabajador al atardecer (Figura 6), y (3) *smartphone* con soporte manual (Figura 7). El vídeo (1) contiene una amplia variedad de puntos de vista de un grupo de 10 trabajadores en una obra de cimentación. Además, las imágenes de las Figs. 5(a), 5(b) y 5(d) presentan importantes oclusiones debidas principalmente a la perspectiva de captura y a la presencia de barras de hierro verticales.

La metodología permite descartar esqueletos de trabajadores afectados por oclusión parcial mediante los umbrales de confiabilidad. Un ajuste típico de $u_{Conf}^E = 0,35$ permite descartar desde el inicio esqueletos de trabajadores con aproximadamente la mitad del cuerpo ocluido. De forma similar, el efecto de la distancia o de las oclusiones parciales en un objetivo pueden controlarse mediante su umbral de tamaño ($u_{Tam}^E = 0,2$).

OP18 está diseñada para inferir puntos antropométricos ocluidos a partir de otros visibles, teniendo menor confiabilidad los primeros. Esta capacidad de inferencia permite a la metodología obtener ROI apropiadas para realizar la clasificación de uso de EPP con oclusiones parciales. Por ejemplo, un $u_{Conf}^{ca} > 0,3$ permite rechazar el análisis del casco del trabajador E1 en la Figura 5(b), mientras que la inferencia lograda con el mismo u_{Conf}^{ca} en su oído y hombro derechos en la Figura 5(e) permite obtener una $E.R^{ca}$ adecuada para clasificar con acierto el uso de casco. Sin embargo, dicho umbral descarta el análisis de los trabajadores E2, E3 y E5 en esta figura.

La importancia de descartar las detecciones con el umbral de tamaño queda clarificada en la Figura 5(f), donde se descartan dos falsas detecciones (D1 y D2). Sin embargo, el método falla en situaciones como la mostrada en la Figura 5(c), donde el esqueleto del trabajador E7 es detectado de forma errónea. En forma resumida, en la Figura 5(a) se detectan de forma correcta E2, E3, E7 y E10 en la Figura 5(b) E1 y E2, en la Figura 5(c) se detecta de forma correcta los chalecos de E7 y E1 y de forma errónea el casco de E7, en la Figura 5(d) E1 y E5, en la Figura 5(e) E1 y por ultimo en la Figura 5(f) E1, E2, E4 y E5 y además se descartan las falsas detecciones D1 y D2.

Las condiciones de captura del segundo vídeo generan muchos cuadros borrosos debidos al movimiento de la cámara (Figs. 6(e) y 6(f)). Sin embargo, la metodología propuesta es capaz de recabar información de utilidad al trabajar solamente

con esqueletos y ROI que poseen mayor confiabilidad. Además, en este vídeo se presentan algunas dificultades mencionadas en la Sección 1, como el cambio gradual de iluminación (Figura 6(f)), las oclusiones inherentes al tipo de trabajo (Figuras. 6(b) y 6(d)) y la diversidad de posturas y tamaños (Figura 6(c)). En las Figuras. 6(d) y 6(e) se ejemplifican en la clasificación de casco de las dos dificultades más comunes halladas en la metodología. En este caso se relaciona con la anormal utilización de una prenda extra en la cabeza, sumado al antes mencionado aspecto borroso generado por el movimiento y a la evidente distancia de la cámara. Sin embargo, en la Figuras. 6(a) el mismo casco del trabajador en primer plano es clasificado de una forma correcta. Finalmente, en la Figura 7 se presenta una secuencia de trabajo en grupo en la que ningún trabajador utiliza EPP. El método propuesto logra discernir correctamente el no uso de EPP en trabajadores con posturas variables, ocluidas, y con distinta indumentaria laboral. Ante todos estos casos, la metodología se muestra robusta, aportando resultados confiables de muy alta precisión (por encima del 99 %).

4. Conclusiones y trabajo futuro

Se propuso un enfoque novedoso para la detección automática de EPP de seguridad en entornos reales de construcción utilizando técnicas de CV y aprendizaje automático. El sistema consta de dos bloques principales: un detector visual de puntos antropométricos del esqueleto humano para localizar personas y un clasificador específico para detectar la ausencia o presencia de EPP en zonas extraídas a partir de dichos esqueletos. El sistema puede reducir las necesidades de supervisión y monitoreo manual, y permitir el control y la recuperación de datos. La metodología propuesta presenta ventajas significativas al ser robusta frente a las frecuentes condiciones que dificultan la correcta detección de EPP a partir de vídeos (orientación, color, pose, oclusiones, movimiento). Las pruebas presentadas se centran en métodos modernos de captura como drones o *smartphones*, que suelen presentar una alta inestabilidad. Además, al estar basada en la detección de esqueletos, la metodología aborda la problemática de una forma innovadora, de la cual no se tiene evidencia previa en el estado del arte. Por todo ello, esta propuesta se presenta como una alternativa de solución válida y práctica, aplicable en condiciones reales de trabajo y adecuada para un monitoreo continuo no supervisado en tiempo real. Como extensión de este trabajo, nuestro futuro objetivo es hacer que el sistema sea escalable e incluir otros EPP como guantes y calzado de seguridad, así como diseñar o evaluar y comparar otras redes para la detección de EPPs.

Agradecimientos

Este trabajo ha sido realizado parcialmente gracias al apoyo del Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina (CONICET), la Junta de Extremadura (España) a través del Fondo Europeo de Desarrollo Regional (código GR18135) y la Universidad Nacional del Sur (código 24/K083).

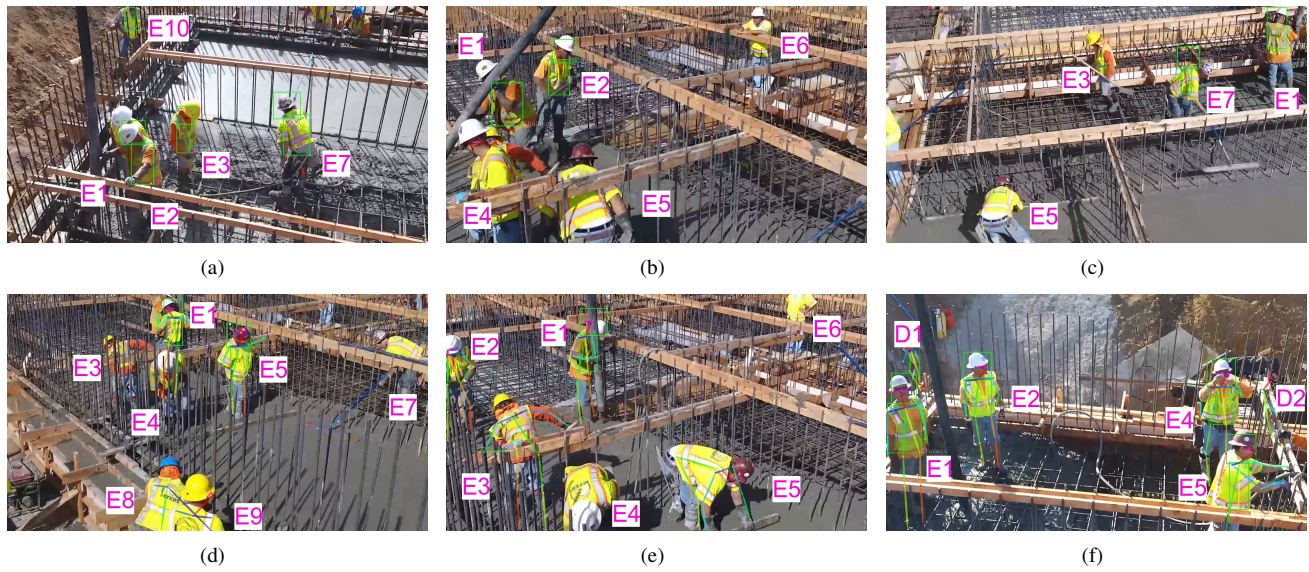


Figura 5: Resultados del vídeo de evaluación 1 (dron).



Figura 6: Resultados del vídeo de evaluación 2 (atardecer).

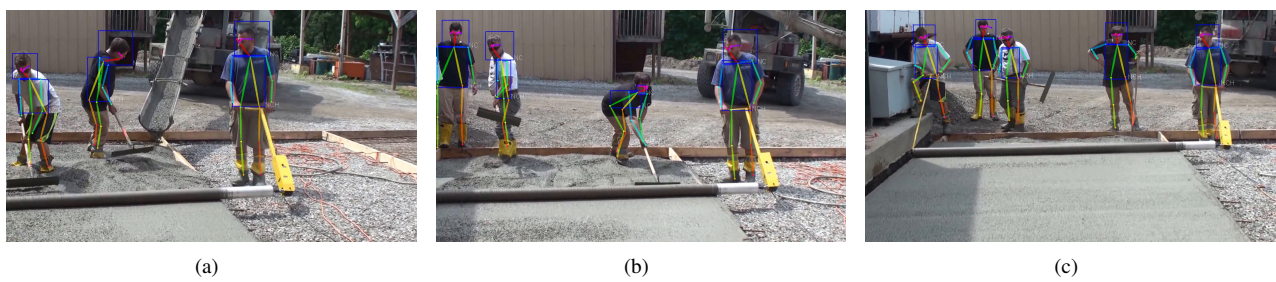


Figura 7: Resultados del vídeo de evaluación 3 (smartphone).

Referencias

- Ankrum, D. R., Nemeth, K. J., 2000. Head and neck posture at computer workstations – what is neutral? In: Proc Human Factors and Ergonomics Soc Annual Meeting. Vol. 44. pp. 565–568.
DOI: 10.1177/154193120004403046
- Arias Gallegos, W. L., 2011. Uso y desuso de los equipos de protección personal en trabajadores de construcción. *Ciencia & Trabajo* 40, 119–124.
- Brilakis, I., Park, M. W., Jog, G., 2011. Automated vision tracking of project related entities. *Advanced Engineering Informatics* 25 (4), 713–724.
DOI: 10.1016/j.aei.2011.01.003
- Cao, Z., Simon, T., Wei, S.-E., Sheikh, Y., 2017. Realtime multi-person 2d pose estimation using part affinity fields. In: Proc IEEE Conf CVPR. pp. 7291–7299.
URL: <https://arxiv.org/abs/1611.08050v2>
- Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T. M., An, W., 2018. Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Automation in Construction* 85 (1), 1–9.
DOI: 10.1016/j.autcon.2017.09.018
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *Proc IEEE Conf CVPR*, 770–778.
DOI: 10.1109/CVPR.2016.90
- ILO, 2002. ILO-OSH 2001. Directrices relativas a los sistemas de gestión de la seguridad y la salud en el trabajo. Oficina Internacional del Trabajo, Ginebra (Suiza).
- Konstantinou, E., Lasenby, J., Brilakis, I., 2019. Adaptive computer vision-based 2d tracking of workers in complex environments. *Automation in Construction* 103, 168 – 184.
DOI: 10.1016/j.autcon.2019.01.018
- Massiris Fernández, M., Delrieux, C., Fernández Muñoz, J. Á., 2018. Detección de equipos de protección personal mediante la red neuronal convolucional Yolo. In: *Actas de las XXXIX Jornadas de Automática*. pp. 1022–1029.
URL: <http://hdl.handle.net/10662/8846>
- Memarzadeh, M., Golparvar-Fard, M., Niebles, J. C., jul 2013. Automated 2D detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors. *Automation in Construction* 32, 24–37.
DOI: 10.1016/J.AUTCON.2012.12.002
- Mneymneh, B. E., Abbas, M., Khoury, H., 2017. Automated hardhat detection for construction safety applications. *Procedia Engineering* 196, 895 – 902.
DOI: <https://doi.org/10.1016/j.proeng.2017.08.022>
- Mosberger, R., Andreasson, H., Lilienthal, A. J., 2014. A customized vision system for tracking humans wearing reflective safety clothing from industrial vehicles and machinery. *Sensors (Switzerland)* 14 (10), 17952–17980.
DOI: 10.3390/s141017952
- Park, J., Yang, X., Cho, Y. K., Seo, J., 2017. Improving dynamic proximity sensing and processing for smart work-zone safety. *Automation in Construction* 84, 111 – 120.
DOI: 10.1016/j.autcon.2017.08.025
- Park, M. W., Brilakis, I., 2012. Construction worker detection in video frames for initializing vision trackers. *Automation in Construction* 28, 15–25.
DOI: 10.1016/j.autcon.2012.06.001
- Park, M.-W., Brilakis, I., 2016. Continuous localization of construction workers via integration of detection and tracking. *Automation in Construction* 72, 129 – 142.
DOI: 10.1016/j.autcon.2016.08.039
- Park, M.-W., Elsafty, N., Zhu, Z., 2015. Hardhat-wearing detection for enhancing on-site safety of construction workers. *Journal of Construction Engineering and Management* 141 (9), 4015024.
DOI: 10.1061/(ASCE)CO.1943-7862.0000974
- Seong, H., Choi, H., Cho, H., Lee, S., Son, H., Kim, C., 2017. Vision-based safety vest detection in a construction scene. In: *ISARC 2017 - Proceedings of the 34th International Symposium on Automation and Robotics in Construction*. Isarc. pp. 288–293.
DOI: 10.22260/isarc2017/0039
- Shrestha, K., Shrestha, P. P., Bajracharya, D., Yfantis, E. A., 2015. Hard-hat detection for construction safety visualization. *Journal of Construction Engineering* 2015.
DOI: 10.1155/2015/721380
- Son, H., Choi, H., Seong, H., Kim, C., 2019. Detection of construction workers under varying poses and changing background in image sequences via very deep residual networks. *Automation in Construction* 99, 27–38.
DOI: 10.1016/j.autcon.2018.11.033
- Unión Europea, P. y. C. d. l. U., 1989. Directiva 89/656/cee relativa a las disposiciones mínimas de seguridad y de salud para la utilización por los trabajadores en el trabajo de equipos de protección individual.
- Xie, Z., Liu, H., Li, Z., He, Y., 2018. A convolutional neural network based approach towards real-time hard hat detection. *Proceedings of the 2018 IEEE International Conference on Progress in Informatics and Computing, PIC 2018*, 430–434.
DOI: 10.1109/PIC.2018.8706269